

# Weekly Report

Lu Junhua

2015 年 7 月 5 日

This week, Ke Junming did many researches on the gongan data, and wrote a brief report. However, I made a mistake on the data samples. 86 people exist in both samples. Although I used “not exist” queries, some duplications remained in the datasets while extraction. This was caused by my unfamiliar of PL/SQL, and I restarted the arrangement of data.

Internet cafe data can be used after discussion with Gu Tianyu, since he made a mistake in configuration of data, and now it can be added as another attribute.

We may also add 2 attributes: number of commit crimes, relatives’ crime statistics. The former one is easy to get, but the later one needs specialized treats. I need more inspection on the relations between 户主 and other people under the same 户口ID.

After Ma Yuxin’s suggestion, I installed scikit learn library in python, and I learned a little bit python and numpy for processing the data. Ke said that SVM in prime-matlab is not so splendid for classification, so I will implement it in python.

On Wednesday, I made a report on Markovian Models, including Markov chains and Hidden Markov Models with three algorithms.

Also this week, I spent much time on graduation-related issues.

---

Next Monday, we will have a discussion with TGRAM company leader and gongan data related employee. We must configure much of the problems in data. For Prof. Gu Wentao’s model, we must use the data in a form which meets his requirement, and for the moment we can’t.

Also, I will try my best to re-extraction the data, and if possible apply `svm.SVC()` to fit a model. Also, cross-validation is necessary for testing the performance of the model.